DOT/FAA/AR-96/67

Office of Aviation Research
Washington, DC 20591

Test and Evaluation Plan of Computer-Based Training for the CTX 5000 Explosives Detection System

S. Cormier, Ph.D.
J. L.. Fobes, Ph.D.

Aviation Security Human Factors
Program, AAR-510
William J. Hughes Technical Center
Atlantic City International Airport, NJ 08405

September 1996

Final

U.S. Department of Transportation
William J. Hughes Technical Center

NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents or use thereof. The United States Government does not endorse products or manufacturers. Trade or manufacturer's names appear herein solely because they are considered essential to the objective of this report.

Technical Report Documentation Page

1. Report No. DOT/FAA/AR-96/67	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle TEST AND EVALUATION PLAN OF COMPUTER-BASED TRAINING FOR THE CTX EXPLOSIVES DETECTION SYSTEM		5. Report Date September 1996 6. Performing Organization Code AAR-510
7. Author(s) S. Cormier, Ph.D. and J. L. Fobes, Ph.D.		8. Performing Organization Report No. DOT/FAA/AR-96/67
9. Performing Organization Name and Address Federal Aviation Administration William J. Hughes Technical Center Atlantic City International Airport, NJ 08405		10. Work Unit No. (TRAIS) 11. Contract or Grant No. DTFA03-94-C-00012
12. Sponsoring Agency Name and Address U.S. Department of Transportation, Federal Aviation Administration Associate Administrator of Civil Aviation Security 800 Independence Ave., S.W. Washington, DC 20591		13. Type of Report and Period Final Report 14. Sponsoring Agency Code ACS-1

15. Supplementary Notes
Prepared by NYMA Inc., William Maguire and Jean Dunn, 500 Scarborough Drive, Egg Harbor Township, NJ

16. Abstract
This Test and Evaluation Plan evaluates the effectiveness of computer-based training (CBT) as an element of a Screener Proficiency Evaluation and Reporting System (SPEARS) for checked baggage screening with the CTX 5000.

The CTX 5000 combines computed tomography imaging and automated detection of explosives. Alarm resolution with this complex system requires that screeners learn to skillfully discriminate system false alarms from system-identified true threats.

Testing is designed to determine the effectiveness of the CBT to meet the critical operational and technical issues described in this plan. The testing will be conducted at airports where the CTX 5000 system is currently located.

17. Key Words Computer-Based Training Computed Tomography Explosive Detection Systems	18. Distribution Statement This document is available to the public through the National Technical Information Service, Springfield, Virginia, 22161

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 58	22. Price

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

PREFACE

This test plan defines the investigation of the effectiveness of computer-baed training (CBT) as an element og a Screener Proficiency Evaluation and Reporting System (SPEARS) for checked baggage screening with the CTX 5000. The key FAA personnel supporting this testing are J. L. Fobes, Ph.D.; E. C. Neiderman, Ph.D.; S. Cormier, Ph.D.; J. M. Barrientos; and B. A. Klock with the Aviation Security Research and Development Division, Human Factors Program (AAR-510).

TABLE OF CONTENTS

APPENDICES

LIST OF ILLUSTRATIONS

ACRONYMS AND ABBREVIATIONS

AvSec HF	Aviation Security Human Factors
ATL	Atlanta International Airport
CBT	Computer-Based Training
COIC	Critical Operational Issues and Criteria
CT	Computed Tomography
CTI	Combined Threat Image
DOT	Department of Transportation
FAA	Federal Aviation Administration
GAO	General Accounting Office
HFE	Human Factors Engineer
IED	Improvised Explosive Device
MOE	Measure of Effectiveness
MOP	Measure of Performance
NAS	National Airspace System
OIC	Other Issues and Criteria
OT&E	Operational Test and Evaluation
SFO	San Francisco International Airport
SPEARS	Screener Proficiency Evaluation and Reporting System
SPI	Safe Passage International Ltd.
T&E	Test and Evaluation
TER	Test and Evaluation Report
TEP	Test and Evaluation Plan
TIP	Threat Image Projection

1. INTRODUCTION

1.1 GENERAL

The Federal Aviation Administration (FAA), in conjunction with the U.S. aviation industry, is developing new equipment and procedures to improve aviation security in the National Airspace System (NAS). Investigation of human factors is critical to the success of these efforts. The President's Commission on Aviation Security and the General Accounting Office (GAO) recognized this need and have recommended that there be a greater focus on human factors and training to complement advanced technologies.

The Screener Proficiency Evaluation and Reporting System (SPEARS) is being developed to improve and maintain the effectiveness of security screening personnel employed at airports. The SPEARS consists of two components: (a) an offline Computer-Based Training (CBT) system to teach screeners to detect various threat objects, and (b) an online threat image projection (TIP) training and testing program is to be employed at airport security checkpoints. This latter configuration is designed to further develop and maintain threat detection proficiency by insertion of simulated threat images into the normal flow of passenger bag images. The effectiveness of the CBT and TIP components will be addressed during separate test and evaluation (T&E) activities.

InVision's CTX 5000 scanner is a new technology application that combines computed tomography (CT) and automatic detection of explosives. This is a more complex system than baggage screeners have previously used. It demands that security personnel use a new set of skills to accomplish the task of screening for Improvised Explosive Devices (IEDs), including the ability to distinguish system false alarms from real threats. CBT and TIP represent important training variables that need to be evaluated carefully.

This is a Test and Evaluation Plan (TEP) for CBT; TIP will be addressed in a separate TEP. It addresses the Critical Operational Issues and Criteria (COIC) and Other Issues and Criteria (OIC) established by the FAA for the CBT component of SPEARS for IED screening of checked baggage with the CTX 5000 system.

1.2 PURPOSE

This SPEARS Operational Test and Evaluation (OT&E) is being conducted to evaluate the ability of CBT to provide checked baggage screeners with the information and threat resolution skills they need to properly operate the CTX 5000. Maintaining a workforce of adequately trained and performing X-ray screening personnel is critical to the mission of aviation security, both domestically and internationally. This TEP outlines the methods and procedures to be used in ensuring that SPEARS training and evaluation for operators of the CTX 5000 system meet the functional requirements established by the FAA as necessary to produce a capable workforce. Specifically, it will examine whether, after CBT (plus limited supplementary classroom instruction), screeners can detect IEDs and distinguish machine false alarms from actual explosive threats. This evaluation will include the collection and analysis of empirical data at two major U.S. airports, San

Francisco International Airport (SFO) and Atlanta International Airport (ATL). It will focus on the operational and technical capabilities of the CTX 5000 system with operators who have received CBT. The CTX system includes the CTX 5000, the operators, operator training, and all threat resolution protocols.

1.3 SCOPE

The focus of this TEP is to evaluate the degree to which a CBT system increases the ability of screeners to successfully resolve machine-generated threat alarms in checked baggage. Detection rates for IEDs (and explosives) using conventional X-ray screening need to be improved. The use of CT offers a number of potential advantages over X-ray screening. The volume images obtained contain much more information than the X-ray image, allowing objects to be viewed without clutter of overlapping images and with higher contrast. At the same time, the use of substantial computing capacity for CT image reconstruction in these scanners facilitates implementation of computer-aided explosives detection. In computer-aided detection, the machine first analyzes the image for the presence of explosives. The human operator then decides which potential threat objects need additional inspection.

The OT&E of CBT components of CT SPEARS will focus on assessment of these training methods against the COIC and OIC described in this TEP. The OT&E will be conducted at SFO and ATL and other airports where the CTX 5000 is currently located.

1.4 BACKGROUND

1.4.1 SPEARS Program

The SPEARS Program was put into its current form in response to a congressional mandate (Aviation Security Improvement Act of 1990, Public Law 101-604). This act directs the FAA to improve aviation security through the optimization of human factors elements in the U.S. airport security system. The evaluation of screener performance and effectiveness was emphasized to identify potential security improvements. An aviation security Department of Transportation (DOT) task force supported this emphasis by concluding that human performance was the critical element in the screening process.

The mandate directed that screeners be effectively trained to use threat detection equipment properly. The detection of explosive and incendiary devices was identified as critically important because of the potential for significant loss of life and aviation resources.

Safe Passage International (SPI), a company that developed CBT for screening with conventional X-ray, then developed CBT modules for the InVision system and the task of IED detection with CT images.

The Airport Demonstration Project for the CTX system was scheduled to begin before the CBT modules were completely developed. For this reason, the Lawrence Livermore National Laboratories, the FAA, and InVision developed a classroom curriculum for CTX

training. This classroom training was used at SFO and ATL before the introduction of the CBT. Some components of this classroom curriculum will continue after CBT is fully operational in order to communicate sensitive and site specific information.

1.4.2 Improvised Explosive Device Screening with the CTX 5000 System

The InVision CTX 5000 is an X-ray-based scanner that automatically screens for explosives. Baggage is automatically fed into the scanner one at a time. The scanner makes an X-ray and CT examination of the bag and computer software then analyzes the CT slices. If the software detects no threat, the bag is CLEARED and unloaded from the scanner. If a potential explosive threat is detected, the computer activates an alarm. At the workstation, the operator is provided with CT and X-ray images of the bag. They also receive an outline of the region identified as a potential threat and information (e.g., density and mass) about the potential threat object. The operator, following the Alarm Resolution Procedures that are emphasized in training, examines the bag and determines whether the threat is real. If the operator determines that the bag is safe, the bag is CLEARED. If the operator cannot determine that the bag is safe, it is declared SUSPECT and additional security procedures are followed.

The CTX 5000 system is a complex system. It requires that screeners learn to operate the controls and to accurately interpret CT slices of checked baggage. The training component is critical to the success of this system. Cognitive and behavioral psychology provides information about how training should be organized, and this information has been incorporated into the design of the TEP.

1.4.3 Cognitive and Behavioral Analysis of Improvised Explosive Device Screening

Maintaining a high level of vigilance and performance in IED detection presents unique problems. The defining feature is that it is a discrimination task, practiced under vigilance conditions, where the signal to be discriminated almost never occurs. This has a number of predictable effects. Absent a special type of training, the screener has no opportunity to see positive targets and, therefore, learn the critical discrimination. CBT is designed to overcome this problem by teaching screeners the basic screening task before they go into the field and exposing them to a variety of threat object and false alarm images.

1.5 FUNCTIONAL REQUIREMENTS

To justify the increased expense that CT screening of checked baggage represents, the system must be capable of detecting IEDs with a very high sensitivity. It must do so without slowing the normal transport of baggage, increasing baggage delivery delays, or

delaying airline takeoffs. The CTX 5000 system can be used to detect IEDs; however, a high detection rate leads to machine false alarms for non-explosives on a certain number of bags.

For this reason, the system is specifically designed to work with human screeners who will examine X-ray and CT images of each bag that is alarmed. They will then determine which alarms should be CLEARED because the bag contains no threat and which alarms are SUSPECT, requiring closer examination. In order for the system to succeed, the following must be true.

a. Screeners must be able to discriminate machine false alarms from genuine threat objects.

b. Screeners must resolve most alarms in a brief period of time (< 30 seconds).

c. Screeners must acquire the ability to do both in a relatively short time after training.

These three requirements are critical for the system to work in the operational environment. They are the results that must be accomplished by the SPEARS training components for the system to be operationally effective.

1.6 SYSTEM DESCRIPTION

Primary training will be accomplished using the Safe Passage® CBT System for CTX 5000. Some supplementary training will be achieved in classroom settings.

1.6.1 Safe Passage CBT System for CTX 5000

The SPI CTX CBT System will be used to train airport baggage screeners to operate the CTX 5000 system and use it to identify IEDs. The system provides a number of features and benefits, as described by SPI.

a. It trains and tests baggage screener skills and techniques for CTX 5000 alarm resolution.

b. It establishes and maintains job performance standards.

c. It tracks the test performance and training history of each security employee.

d. It provides an image library of threats packed in passenger baggage.

e. It allows management to distribute critical bulletins electronically to security personnel.

f. It prints management reports.

g. It provides a simulated CTX 5000 user interface.

1.6.2 Technical Specifications

The system consists of a combination of equipment and software. The following is a list of the system technical specifications.

a. Apple 1710 AppleVision RGB Monitor.

b. Apple Design Keyboard II and mouse that operators use to interact with the training program.

c. Power Macintosh 7500/100 with 32 meg RAM, 1 gigabyte internal hard drive, and quad speed internal CD-ROM drive.

d. Optionally configurable system for remote network link.

1.6.3 CBT Modules

a. Basic CTX 5000 Screening Modules – These modules train screeners to operate the CTX 5000 system. They explain how the system works and familiarize the screeners with the basics of IEDs. They explain in detail the alarming process and the proper procedures to resolve alarms. Each module provides self-paced lessons and exams.

b. Image Library – The image library is a collection of passenger bag images that contain IEDs within a cluttered bag. Each contains X-ray and CT views and is accompanied by descriptive text explaining their distinguishing characteristics. This module serves as a reference for viewing X-ray images of a variety of explosive devices and innocent travel items.

c. Proficiency Testing – Screeners are tested after each lesson block on the information presented. In addition, following the completion of training, screeners are given a performance test using stored CT images of baggage containing IEDs interspersed with bags that generate machine false alarms.

1.6.4 Classroom Training

Some components of training, because they are sensitive, site specific, or simply more effectively communicated in a lecture setting, are to be given in a classroom. These components will be derived from portions of the CTX 5000 Training Syllabus for Screeners developed by the Lawrence Livermore National Laboratory. Classroom materials include handouts, inert explosive simulants, videos, slides, and a Modular Bomb Set.

1.7 Testing Overview

There are two classes of issues for the assessment of CBT against the functional requirements: COIC and OIC. These issues can involve substantially different requirements and investigation methods. The overall OT&E is organized to evaluate each issue, using the most appropriate method.

1.7.1 Critical Operational Issues and Criteria

The COIC are necessary to evaluate the CBT's operational requirements. Each issue is analyzed in terms of one or more criteria by which the system is judged. Each criterion gives rise to one or more Measures of Performance (MOPs) and Measures of Effectiveness (MOEs). Testing will involve collecting data during actual training and in post-training evaluations to assess the effectiveness of the SPEARS CBT testing and training on screener IED detection performance and CTX system operation. Objective observations and screener feedback will also be collected to determine the usability of the CBT interface and the CBT materials.

1.7.2 Other Issues and Criteria

The OIC are supplementary, more specific, or are technical in nature, such as system customization, screener capabilities reporting, feedback, security, insertion, and image content. The OIC will be investigated using structured protocols during field and laboratory testing of system features. Many of the checklists for these tests use the Human Factors Deficiency Rating Scale, Appendix A.

1.7.3 Test Management

The organization of testing is outlined in Appendix B. The OT&E milestones are listed in Table 1.

TABLE 1. OPERATIONAL TEST AND EVALUATION MILESTONES

MILESTONE	DATE	RESPONSIBLE ORGANIZATION
TEP Finalized	9/96	AvSec HF Program
Operational Testing	10/96	AvSec HF Program
Draft Test & Evaluation Report (TER)	3/97	AvSec HF Program
Final TER	5/97	AvSec HF Program

1.8 Computed Tomography Computer-Based Training Critical Operational Issues and Criteria

1.8.1 Issue 1 - Training Effectiveness

Does CBT enhance screeners' ability to operate the CTX 5000 and resolve alarms?

Criterion 1-1 Following training, screeners operate the system capably.

MOP 1-1-1 Evaluation of screener ability to perform basic operating and safety procedures.

MOE 1-1-1 Screeners make no errors on operating and safety procedures.

Criterion 1-2 Screeners can discriminate IEDs from machine false alarms after training.

MOP 1-2-1 Percentage of screeners passing CBT end-of-course screening test.

MOP 1-2-2 Percentage of correctly identified test items by screener.

MOE 1-2-1 > 90 percent of screener trainees pass test.

Criterion 1-3 Alarm resolution is performed within acceptable time limits.

MOP 1-3-1 Time to resolve true alarms and machine false alarms during the timed final test.

MOE 1-3-1 80 percent of test images are resolved within the required time limits.

MOP-1-3-2 Mean and distribution of time to resolve alarms during first week on system.

MOE 1-3-2 80 percent of alarms are resolved within 20 seconds.

The CBT includes a final test module that presents actual IED and false alarm bag CTIs on a simulated CTX 5000 interface. Both accuracy and time to resolve alarms are measured during final testing (after completion of training) and can be compared to the minimum requirements prescribed. If training is effective, the large majority of screeners should be able to perform within the stated requirements upon first testing.

Cameras with video recording capability will be posted at the test sites. Videotapes of CTX operation will be used to evaluate the ability of newly trained screeners to carry out basic operations. The Basic Operations Checklist (Appendix C) will be used to organize the evaluation.

1.8.2 Issue 2 - Usability of Interface and Training Materials

Are there any CBT interface factors or procedures that degrade training effectiveness?

Criterion 2-1 Investigative in nature.

MOP 2-1-1 Deficiencies noted in CBT usability.

MOP 2-1-2 The frequency and type of screener errors observed due to hardware, software, or procedures.

MOE 2-1-1 CBT does not have serious interface deficiencies.

Are materials interactive and comprehensible?

Criterion 2-2 The training materials have no deficiencies that would seriously hamper their usefulness.

MOP 2-2-1 Deficiencies noted in the training materials.

What level of reading comprehension is required for the written training materials?

Criterion 2-3 The required reading comprehension level is appropriate.

MOP 2-3-1 The CBT reading grade level.

MOE 2-3-1 Reading grade level of lessons does not exceed 8th grade.

Criterion 2-4 All technical and unfamiliar terms are defined before they are used.

MOP 2-4-1 Deficiencies in defining technical terms.

MOE 2-4-1 No undefined technical terms are found in the CBT.

The evaluation of the interface usability will be made both in the pre-CBT evaluation phase and during the observed training. In the pre-training phase, human factors engineers (HFEs) will test the system using an abridged version of the SPEARS HFE Usability Checklist (see Appendix D. This checklist is an adaptation of the *Guidelines for the Design of User Interface Software* (Smith & Mosier, 1986) and MIL-STD-1472D (Department of Defense, 1989). During the observed training, problems that arise in the screeners' interactions with the CBT will be recorded. The issues and concerns raised during an earlier evaluation of the CTX interface (Fobes, Cormier, & Barrientos, 1996), will be considered in evaluating the simulated interface. Immediately following the training, screeners will be asked to complete the Safe Passage Screener Survey (Appendix E). Instructors will be asked to complete the Safe Passage Instructor Survey (Appendix F). Informed consent will be obtained from all screeners before they are given questionnaires (Appendix G).

Training materials will be evaluated by HFEs during the pre-training evaluation and during the observed training period. In addition, screener feedback about the perceived usability of the training will be received from the Safe Passage Screener Survey given to screeners immediately after training. The reading comprehension level will be assessed by calculating the reading grade level for the CBT using the Flesch-Kincaid Grade Level

index within Microsoft Word 6.0 software. This index is calculated by loading five random 150-word samples of text from the CBT into Microsoft Word. The Flesch-Kincaid equation computes grade level based on the average number of words per sentence.

1.9 OTHER ISSUES AND CRITERIA

1.9.1 Issue 3 - Image Content

Do the IED images used for training represent the range of current threats?

Criterion 3-1 The training material will include combined threat images (CTIs) from all the explosive threat categories.

MOP 3-1-1 Numbers of CTIs in each threat category.

MOE 3-1-1 Adequate numbers of CTIs exist in all threat categories.

During the pre-CBT evaluation phase, image content will be evaluated by HFEs.

1.9.2 Issue 4 - Customization

Can training be tailored to individual screener's needs?

Criterion 4-1 Images can be selected for an individual screener as a function of IED type.

MOP 4-1-1 Problems encountered while creating customized image sets.

MOE 4-1-1 Images can be selected by threat category.

Criterion 4-2 Additional images can be introduced into the CBT.

MOP 4-2-1 Problems with insertion of new images in the CBT.

MOE 4-2-1 No severe problems with introduction of new images.

Customization will be examined during the post-training evaluation. Appendix H, CBT Customization Checklist, was adapted from an earlier document (Fobes, McAnulty, & Lofaro, 1995).

1.9.3 Issue 5 - Feedback

Is feedback provided during training?

Criterion 5-1 Feedback is provided for all questions about lesson content and for all evaluations of threats.

MOP 5-1-1 Deficiencies noted in feedback after incorrect answers.

MOE 5-1-1 No severe deficiencies in feedback are noted.

CBT feedback will be examined in the pre-training evaluation and during the observed training. The CBT Feedback Checklist (Fobes, McAnulty, & Lofaro, 1995) will structure these evaluations (Appendix I). It has been modified so that it is appropriate for feedback evaluation of content questions and alarm resolution testing.

1.9.4 Issue 6 - Screener Capability Summaries

Are useful training reports prepared?

Criterion 6-1 Training reports describe a screener's performance, in general and for specific alarm threat categories. Both relative and absolute measures of performance are provided.

MOP 6-1-1 Deficiencies in individual screener reports.

MOE 6-1-1 No severe deficiencies in individual reports.

Criterion 6-2 Cumulative statistics are kept of training performance on the CTX 5000 for all facilities where training takes place and summary statistics are maintained.

MOP 6-2-1 Deficiencies in site summary reports.

MOE 6-2-1 No severe deficiencies noted in the reports or access to them.

These capabilities will be tested in the post-training phase of the CBT evaluation, when there has been training at multiple sites, so that all capabilities can be tested. Appendix J, CBT Capabilities Checklist, is adapted from Fobes, McAnulty, & Lofaro (1995).

1.9.5 Issue 7 - Security

Is access restricted?

Criterion 7-1 Only authorized personnel can access particular aspects of the system.

MOP 7-1-1 Deficiencies noted in system security.

MOE 7-1-1 No severe deficiencies in security.

During the pre-training evaluation, HFEs will test system security. Deficiencies will be noted by use of the CBT Security Access Control Checklist (Appendix K).

2. OPERATIONAL TEST AND EVALUATION

2.1 SUBJECTS

The exact number of subjects will depend upon the availability and suitability of candidate screeners at the airport test sites. Screeners who were initially trained at these sites, and who received little or no CBT, will not be used. Only screeners who have received the full CBT and were not previously exposed to training in the use of the CTX 5000 will be evaluated. We anticipate that the number of subjects will range from 12 to 24.

2.2 EQUIPMENT

CBT will be provided by the SPI System. Associated classroom training will be provided using the lecture syllabus.

2.3 DATA COLLECTION PROCEDURES

2.3.1 Initial Evaluation of Computer-Based Training by Human Factors Engineers

Once CBT is deemed operationally ready by Safe Passage, HFEs will evaluate a number of system features before it is used in the field. These tests will be conducted at a site to be determined and will include the following evaluations:

a. Security evaluation: Using the Security Access Control Checklist (Appendix K), determine how access is controlled for system features (e.g., customized image presentation, reports).

b. Usability and Feedback evaluation: Using the appropriate checklists (Appendices D & I), the usability of the system and the feedback provided to users will receive a preliminary evaluation.

c. Readability evaluation: Samples of text from each lesson are chosen, and readability scores are calculated.

2.3.2 Observed Computer-Based Training

Screeners will receive 4 days training using the CBT modules, additional lecture information, and hands-on training with the equipment. The final lesson of the CBT includes an examination with an IED detection test. IED test performance will be recorded for later analysis.

a. HFEs will observe actual training sessions. All problems with the screeners use of the equipment and their understanding of the materials will be recorded. After recording, they will be evaluated against the basic operations and usability checklists (Appendices C & D).

b. A second source of information about the training will come from the SPI Screener Survey (Appendix E), which will be distributed to all trainees immediately following completion of CBT.

2.3.3 Post-Training Evaluations

The evaluation of customization and report capability (Appendices H & J) requires that some number of individuals have completed the CBT. Therefore, testing of these features will be delayed until more than one group and more than one site have used CBT.

a. HFEs will attempt to produce customized image sets in training modules and tests.

b. Archived reports of screener activities will be examined.

2.3.4 Video Recording of CTX 5000 Screening

The post-training ability of screeners to operate the equipment will be evaluated by gathering and reviewing video records of screeners' ability to perform basic CTX operations. The information gathered will include time to resolve threat alarms that occur.

2.4 DATA ANALYSIS PROCEDURES

CBT effectiveness is simply summarized in terms of number of screeners who meet the accuracy and time requirements. These percentages are then compared to MOEs 1-2-1 and 1-3-1.

The applicable video records will be collected and examined offline at FAA facilities. HFEs will determine whether each screener knows operational and safety procedures for the equipment based upon screeners' activities in the first week of work. Additionally, each screener's time to resolve alarms for five alarms each day will be collected from the video records. This information, plotted over days, will enable us to determine whether screeners' performance is changing during the first few days on the job.

The SPI Screener and Instructor Survey (Appendices E and F) results can be summarized by reporting the means and distributions of the screeners' answers. This information, combined with the recorded deficiencies in usability and feedback recorded by the HFEs in pre-training and during observed training, will form the basis for evaluation of issues 2, 4, and 6.

Other issues and criteria can be evaluated by noting all deficiencies using appropriate checklists and evaluation criteria.

3. REFERENCES

Forbes, J. L., Cormier, S. M., & Barrientos, J. M. (1996). *An evaluation of the explosives detection system CTX 5000 user interface for alarm resolution* (DOT/FAA/AR96/50). Atlantic City, NJ: William J. Hughes Technical Center.

Fobes, J. L., McAnulty, D. M., & Lofaro, R. (1995). *Test and evaluation plan for the screener proficiency evaluation and reporting system (SPEARS) computer-based training* (DOT/FAA/CT-95/10). Atlantic City, NJ: FAA Technical Center.

Smith, S. L., & Mosier, J. N. (1986). *Guidelines for the design of user interface software* (Technical Report ESD-TR-86-278). L. G. Hanscom Air Force Base, MA: U.S. Air Force.

U.S. Department of Defense (1989). *Military standard: Human engineering design criteria for military systems, equipment and facilities* (MIL-STD-1472D). Washington, D.C.

APPENDIX A

HUMAN FACTORS DEFICIENCY RATING SCALE

Severity	Description
Severe	There is a high probability of operational failure, severe damage, loss of equipment, and injury to operators or passengers.
Major	There is a high probability of degraded system performance, major damage to equipment, or discomfort to operators or passengers.
Moderate	There may be no measurable impact on system performance, though there is a measurable impact upon the performance of system components or sub-systems (including the human subsystem). Operators or passengers try to compensate for, or work around, system defects.
Minimal	There is no measurable impact on the performance of system components or subsystems (including the human subsystem), although operators' or passengers' negative attitudes toward features to the system may be measurable.
Negligible	The problem has a negligible impact on short-term system performance. There is no measurable impact on operator or passenger attitudes.
None	No problem or negative factor related to system performance is noted.

APPENDIX B

CBT TESTING DESIGN

PROCEDURE	Pre-Training Evaluation	Observed Training	Field Testing	Evaluation of CTX Operation Capabilities
ACTIVITY	Test and preview: CBT security features, usability of materials, and interface. Readability of lessons calculated.	12 -24 screeners receive CBT. Usability and feedback capabilities of CBT evaluated by HFE during training. Screeners evaluate CBT using Safe Passage Screener Survey. Instructors evaluate CBT using Safe Passage Instructor Survey.	Test CBT for Report Capabilities and Customization.	HFEs review video records and evaluate operational capabilities of newly trained screeners.
DATA COLLECTED	Kind and magnitude of security and usability deficiencies. Flesch-Kincaid Grade Level Indices.	Frequency distributions of test scores. Kind and magnitude of usability and feedback deficiencies. Frequency distributions of responses to Safe Passage Screener Survey and Safe Passage Instructor Survey.	Kind and magnitude of deficiencies in customization and capabilities reporting.	Kind and magnitude of deficiencies in screeners operational capabilities. Frequency distribution of threat resolution times by days.
DURATION	2 DAYS	4 DAYS for each group trained.	2 DAYS	1 WEEK for each group trained.

APPENDIX C

BASIC OPERATIONS CHECKLIST

Activities	Deficiency Rating	Comments
Can bring the system to operational readiness.		
Can turn on the system.		
Can bring system back from a fault.		
Can handle diebacks and reset failures.		
Follows proper security procedures.		
Uses scanning modes correctly.		
Can reset scanning modes.		
Can handle routine and emergency shutdowns.		
Follows proper safety procedures.		
Uses the Disable X-ray mode properly.		
Uses proper procedures to clear baggage jams.		
Performs daily shutdown properly.		
Handles alarms properly.		
Can operate trackball.		
Uses threat resolution tools.		
Can properly request more slices.		
Knows procedures for unresolved alarms and potential hazards.		
Resolves threats promptly.		
Recorded threat resolution times.		

APPENDIX D

CBT USABILITY CHECKLIST

Human Factors Principle	Deficiency Rating	Comments
DATA ENTRY		
1. Users need enter data only once.		
2. Display feedback for all user actions during data entry; display keyed entries stroke by stroke.		
3. Provide fast response by the computer in acknowledging data entries.		
4. Incorporate a consistent method for data change.		
5. When critical data are to be processed, require an explicit "Enter" action to initiate the processing.		
6. Provide feedback for the completion of data entry.		
7. Make field labels consistent; always employ the same label to indicate the same kind of data.		
DATA DISPLAY		
1. Ensure that whatever data a user needs for any transaction will be available for display.		
2. Do not overload displays with extraneous data.		
3. For any particular type of data display, maintain consistent format from one display to another.		
4. Ensure that each data display will provide needed context, recapitulating prior data as necessary so that a user does not have to rely on memory to interpret new data.		

Human Factors Principle	Deficiency Rating	Comments
5. The wording of displayed data and labels should incorporate familiar terms and the task-oriented jargon of the users.		
6. Choose words carefully and then use them consistently.		
7. When abbreviations are used, choose those abbreviations that are commonly recognized and do not abbreviate words that produce uncommon or ambiguous abbreviations.		
8. Ensure that abbreviations are distinctive so that abbreviations for different words are distinguishable.		
9. When a critical passage merits emphasis to set it apart from other text, highlight that passage by bolding, brightening, color coding, or some auxiliary annotation.		
10. Organize data in some recognizable order to facilitate scanning and assimilation.		
11. In designing text displays, especially text composed for user guidance, strive for simplicity and clarity of wording.		
12. Use consistent logic in the design of graphic displays and maintain standard format, labeling, etc.		
13. Tailor graphic displays to user needs and provide only those data necessary for user tasks.		
14. When graphics contain outstanding or discrepant features that merit attention by a user, consider displaying supplementary text to emphasize that feature.		

Human Factors Principle	Deficiency Rating	Comments
15. When a user's attention must be directed to a portion of a display showing critical or abnormal data, highlight that feature with some distinctive means of coding.		
16. Adopt a consistent organization for the location of various display features from one display to another.		
17. Assign consistent meanings to symbols and other codes, from one display to another.		
18. Choose colors for coding based on conventional associations with particular colors.		
SEQUENCE CONTROL		
1. Defer computer processing until an explicit user action has been taken.		
2. Employ similar means to accomplish ends that are similar, from one transaction to the next, from one task to another, throughout the user interface.		
3. Display some continuous indication of current context for reference by the user.		
4. Adopt consistent terminology for online guidance and other messages to users.		
5. Choose names that are semantically congruent with natural usage, especially for paired opposites (e.g., UP/DOWN).		
6. Ensure that the computer acknowledges every entry immediately; for every action by the user there should be some apparent reaction from the computer.		

Human Factors Principle	Deficiency Rating	Comments
7. When a user is performing an operation on some selected display item, highlight that item.		
8. Design the interface software to deal appropriately with all possible control entries, correct and incorrect.		
9. When a user completes correction of an error, require the user to take an explicit action to reenter the corrected material; use the same action for reentry that was used for the original entry.		
10. When a control entry will cause any extensive change in stored data, procedures, and/or system operation, and particularly if that change cannot be easily reversed, notify the user and require confirmation of the action before implementing it. Provide a prompt to confirm actions that will lead to possible data loss.		
USER GUIDANCE		
1. When the computer detects an entry error, display an error message to the user stating what is wrong and what can be done about it.		
2. Make the wording of error messages as specific as possible.		
3. Make error messages brief but informative.		
4. Adopt neutral wording for error messages; do not imply blame to the user, or personalize the computer, or attempt to make a message humorous.		
5. The computer should display an error message only after a user has completed an entry.		

Human Factors Principle	Deficiency Rating	Comments
6. Provide reference material describing system capabilities and procedures available to users for online display.		
7. Along with explicit and implicit aids, permit users to obtain further online guidance by requesting HELP.		
DATA TRANSMISSION		
1. Choose functional wording for terms used in data transmission, including messages, for initiating and controlling message transmission and other forms of data transfer, and for receiving messages.		
2. Design the data transmission procedures to minimize memory load on the user.		
3. Design the data transmission procedures to minimize required user actions.		
DATA PROTECTION		
1. Provide automatic measures to minimize data loss from computer failure.		
2. Protect data from inadvertent loss caused by the actions of other users.		
3. Provide clear and consistent procedures for different types of transactions, particularly those involving data entry, change and deletion, and error correction.		
4. Ensure that the ease of user actions will match desired ends; make frequent or urgent actions easy to take, but make potentially destructive actions sufficiently difficult that they will require extra user attention.		

Human Factors Principle	Deficiency Rating	Comments
5. When displayed data are classified for security purposes, include a prominent indication of security classification in each display.		
6. When a user requests LOG OFF, check pending transactions involving data entry/change and, if data loss seems probable, display an appropriate advisory message to the user.		
VISUAL DISPLAYS		
1. Sufficient contrast shall be provided between displayed information and the display background to ensure that the required information can be perceived by the operator under all expected lighting conditions.		
2. Displays shall be located and designed so that they may be read to the degree of accuracy required by personnel in normal operating or servicing positions without requiring the operator to assume an uncomfortable, awkward, or unsafe position.		
3. Where alphanumeric characters appear on CRT-like displays, the font style shall allow discrimination of similar characters, such as the letter l and the number 1 and the letter z and the number 2.		

AUDIO DISPLAYS		
1. The voice used in recording verbal signals shall be distinctive and mature.		
2. Verbal signal shall be presented in a formal, impersonal manner.		
3. In selecting words to be used in audio warning signals, priority shall be given to intelligibility, aptness, and conciseness in that order.		
4. The volume (loudness) of an audio warning signal shall be designated to be controlled by the operator, the sensing mechanism, or both.		

APPENDIX E

SAFE PASSAGE SCREENER SURVEY

SAFE PASSAGE SCREENER SURVEY

SUBJECT NUMBER: _____ DATE: _____

The FAA wants to know how useful this training is and how it can be made better. Your opinion will be important in improving the Safe Passage system. Your responses will be kept secret, so please give ratings that are your honest opinion of the training. We thank you for your help.

If you have any questions while taking this survey, please ask the FAA representative. This survey has two rating sections, Training and Usefulness. The FAA representative will assist you with each section of the survey.

I. TRAINING

The numbered sentences on the next page refer to the training that you have just received. Read each sentence and decide whether you agree with it. Use the rating scale to express your agreement or disagreement. The scale goes from 1 to 5. If you circle 5, it means you agree very much with the sentence. If you circle 1, it means you do not agree at all. Use numbers in the middle to express agreement somewhere between the extremes. Circle one number for each sentence.

SAFE PASSAGE SCREENER SURVEY

SUBJECT NUMBER: _____ DATE: _____

	Not at all				Very much
	1	**2**	**3**	**4**	**5**
1. The training improved my ability to detect IEDs.	**1**	**2**	**3**	**4**	**5**
2. The training will help me on my job.	**1**	**2**	**3**	**4**	**5**
3. I feel I can operate the CTX equipment skillfully since taking the training.	**1**	**2**	**3**	**4**	**5**
4. I enjoyed taking the Safe Passage training.	**1**	**2**	**3**	**4**	**5**
5. The training system was easy to use.	**1**	**2**	**3**	**4**	**5**
6. I was very tired at the end of the training.	**1**	**2**	**3**	**4**	**5**
7. The image library helped me to better detect different configurations of IEDs.	**1**	**2**	**3**	**4**	**5**
8. The tests at the end of the lessons helped me to remember the training material.	**1**	**2**	**3**	**4**	**5**
9. The tests in the Basic Screening Testing section helped me remember information from the training courses.	**1**	**2**	**3**	**4**	**5**
10. The training helped me learn the procedures for screening passenger bags.	**1**	**2**	**3**	**4**	**5**
11. The certification tests were easy after taking the training.	**1**	**2**	**3**	**4**	**5**
12. The certification tests were hard because I could not remember the information presented in the training lessons.	**1**	**2**	**3**	**4**	**5**

SAFE PASSAGE SCREENER SURVEY

SUBJECT NUMBER: _____ DATE: _____

II. USEFULNESS

Please rate the lessons in terms of how useful they were. The rating scale goes from 1 to 5, with 1 being the least useful and 5 being the most useful.

Circle one number for each training component.

	Not useful				Very useful
	1	2	3	4	5
Overview of the CTX 5000	1	2	3	4	5
Basic X-ray Theory	1	2	3	4	5
X-ray Interpretation	1	2	3	4	5
Transition from X-ray to CT	1	2	3	4	5
Understanding Slices and Shapes	1	2	3	4	5
Explosives and their Components	1	2	3	4	5
CT Interpretation	1	2	3	4	5
Threat Resolution	1	2	3	4	5
Monitoring and Troubleshooting	1	2	3	4	5
Classroom Presentations	1	2	3	4	5
Training on the CTX 5000	1	2	3	4	5
Lesson by Lesson Tests	1	2	3	4	5
Alarm Resolution Testing	1	2	3	4	5

APPENDIX F

SAFE PASSAGE INSTRUCTOR SURVEY

SAFE PASSAGE INSTRUCTOR SURVEY

SUBJECT NUMBER: _____ DATE: _____

The FAA wants to know your opinion of the Safe Passage training. Your ratings will play an important role in evaluating the Safe Passage system. The success of this evaluation relies on your opinions of the training. Your participation is greatly appreciated.

This survey is divided into two rating section: Training and Usefulness.

I. TRAINING

Please circle the number that best indicates your agreement with the statements in the following survey.

	Not at all				Very much
	1	**2**	**3**	**4**	**5**
1. The training improved the trainees' ability to detect IEDs.					
2. The trainees enjoyed taking the Safe Passage training.					
3. The training system was easy for them to use.	**1**	**2**	**3**	**4**	**5**
4. They appeared fatigued at the end of the training.	**1**	**2**	**3**	**4**	**5**
5. The image library contained a good variety of configurations of IEDs.	**1**	**2**	**3**	**4**	**5**
6. The tests at the end of the lessons adequately tested their knowledge of the lesson material.	**1**	**2**	**3**	**4**	**5**
7. The tests in the Basic Screening Testing section were a fair test of the course content.	**1**	**2**	**3**	**4**	**5**

SAFE PASSAGE INSTRUCTOR SURVEY

SUBJECT NUMBER: _____ DATE: _____

	Not at all				Very much
	1	2	3	4	5
8. Following training, the screeners were well acquainted with the procedures for screening passenger bags.	1	2	3	4	5
9. The certification tests were geared to a proper difficulty level.	1	2	3	4	5
10. The lessons were understandable and did not require frequent explanations.	1	2	3	4	5
11. The lessons were complete and did not require a substantial amount of supplemental instruction.	1	2	3	4	5

SAFE PASSAGE INSTRUCTOR SURVEY

SUBJECT NUMBER: _____ DATE: _____

II. USEFULNESS

Please rate the lessons in terms of how useful they were. The rating scale goes from 1 to 5, with 1 being the least useful and 5 being the most useful.

Circle one number for each training course component.

	Not useful				Very useful
	1	**2**	**3**	**4**	**5**
Overview of the CTX 5000	1	2	3	4	5
Basic X-ray Theory	1	2	3	4	5
X-ray Interpretation	1	2	3	4	5
Transition from X-ray to CT	1	2	3	4	5
Understanding Slices & Shapes	1	2	3	4	5
Explosives and their Components	1	2	3	4	5
CT Interpretation	1	2	3	4	5
Threat Resolution	1	2	3	4	5
Monitoring & Troubleshooting	1	2	3	4	5
Classroom Presentations	1	2	3	4	5
Training on the CTX 5000	1	2	3	4	5
Lesson by Lesson Tests	1	2	3	4	5
Alarm Resolution Testing	1	2	3	4	5

APPENDIX G

INFORMED CONSENT AND PERSONAL INFORMATION FORMS